Efficient Visual Saliency Detection in Video Based on Trajectory Clustering

Man Hua[1, a], Ruichun Lin [1,b]

[1]School of Computer Science , Civil Aviation Flight University of China, GuangHan, Sichuan,618307, China

[a]hua.man@163.com, [b]lrc_beat@hotmail.com

Abstract. In this paper, we propose an efficient visual saliency detection method based on trajectory clustering. We group the corner point trajectories using a two stage clustering algorithm. The most stable trajectories are pre-clustered using mean shift in the first stage. Then, we proposed an unsupervised clustering method to cluster the trajectories and detect the number of motions automatically. At last, the motion saliency map is generated with the segmented spare feature points. Experimental results on different videos demonstrate the utility and performance of the proposed approach.

Introduction

A rapid increase has taken place in the usage of multi-media information over recent years. Moving object detection and segmentation is an important research topic in the field of computer vision in multi-media information processing, which is widely used in the area of video surveillance and video compression. At present, there are 3 methods for motion detection: optical flow method, frame difference method and background subtraction method. Because of the detecting speed and effects are more ideal of background subtraction method, it attracted a lot of researcher's interest in recent years. The main idea of background subtraction is use the background subtraction frame subtracting background difference, and use the threshold to get the two values of difference. As a result, we get the moving target template. Therefore, the effective background modeling and the setting of threshold are the key points of background subtraction.

Visual saliency detection is a kind of visual attention mechanism of human detection method in video based on region of interest. In recent years, the visual saliency detection attracted more and more researchers in the field of computer science and psychology of attention which can be applied to target recognition, video coding and event detection. For example, Itti et al [1] apply the detection of visual saliency in video coding, through the visual mechanism simulation, computational resources is priority allocated to those regions which is easy to cause the viewer's attention , which can greatly improve the efficiency of the existing video analysis. In document [2], visual attention model is applied to image retrieval, and obtained good results. These applications are based on the Visual Saliency Detection. Therefore, how to detect the visual attention region more effectively is an important issue in the current study.

In this paper, using a segmentation method based on trajectory clustering on account of the track is formed by the matching of feature points and the trajectory clustering can be better applied to rigid and non-rigid motion. In addition, different object trajectory have big difference in length and direction, so it can be used to distinguish between multiple moving objects. Finally, using the

trajectory clustering can also distinguish the global motion caused by camera moving and local motion of an object.

Aiming at the characteristics of motion in visual saliency region and the problems in existing methods, we proposed a novel segmentation method for visual saliency detection. The method using a hierarchical clustering method to cluster trajectories of feature points according to the characteristics of visual saliency detection. Firstly, classify trajectories in the time domain according to the length of them; secondly, using an unsupervised clustering algorithm for different types of motion segmentation and obtain the motion classification number automatically; finally, using the motion segmentation results , we proposed a method combining spatial and color sampling to generate the motion saliency region.

Motion segmentation based on trajectory clustering

This paper uses the Harris corner-point detection algorithm [3] to detect the feature points in video frames, and with the Pyramid-based feature detection and tracking method [4] to estimate the position of the feature points in the next frame. Assuming that a frame of image have N feature points, in the time of t the i feature point's descriptor is: $p_i^t(x_i^t, y_i^t), i \in N$, where x_i^t, y_i^t is the position of feature point and $x_i^{'}, y_i^{'}$ is the position of this point in the next frame, dx_i, dy_i is motion vector of feature point , where $dx_i = x_i - x_i^{'}, dy_i = y_i - y_i^{'}$, N is the number of feature points. First, remove the feature points which have too large motion vectors: the motion vector of feature points is greater than a threshold would be regarded as production of feature point matching error, and remove it. Second, obtain the M frames' trajectories of each feature point by using the algorithm based on path coherence be proposed in [5]. In this paper, $x_i = \{p_i^1, p_i^2 ... p_i^M\}$ is used to represent trajectory of the i feature point.

In [6], the paper using the expectation maximization algorithm through iterative method to obtain trajectory classification, but in advance, the total number of motion model should be determined when using the EM method. In [7], trajectory also be used in motion segmentation, but the background of this method is static and the length of trajectory is fixed. Nevertheless, in the condition of camera motion will produce a large number of incomplete trajectory. Therefore, the method in [7] couldn't be directly clustering. Studies show in [8], better result can be obtain if clustering of similar length trajectory. Consequently, trajectories are divided into two categories according to the duration of time, namely the complete trajectory and incomplete trajectory. The complete trajectory of feature points are visible in all frames. While the initial frames of incomplete trajectory and the end of it may not be the same. Thus, the similarity comparison of incomplete trajectory is extremely difficult.

Based on the complete trajectory, this paper presents a trajectory clustering algorithm according to the motion information of trajectory, and the segmentation information of feature point contains in trajectories can be obtained after the trajectory clustering results be got.

Generation of the motion Saliency

This paper proposed a motion saliency map generation method combine two attributes of color and space at feature point because of the sparse feature of feature point. Firstly, we obtain the

main clustering color of each feature point, calculate the smallest rectangular area surrounded by the feature points of each cluster and get the color histogram H of it. Determine the maximum peak position P_{mbin} and keep the eight connected bin of P_{mbin} in H, remove the other bins. Using H as color feature of the feature points set.

Secondly, we divide the current image frame into multiple plurality of 8X8 block size, extract the color histogram as color feature of this block. Suppose a feature point clustering contains NS feature points, then the similarity S_i between the i block and the feature point set of it is obtained by equation (1):

$$S_i = \frac{1}{NS} \sum_{j=1}^{NS} \exp\left[-\frac{dd(i,j)^2}{2a^2} - \frac{di(i,j)^2}{2b^2} \right] \tag{1}$$

Spatial distance and color distance of a block to a feature point clustering were calculated by equation (1), where $dd(i,j)$ indicate the spatial distance from i block to j block, $di(i,j)$ is histogram intersection from i block to j block, a and b is the adjustment coefficient of spatial and color distance and set to 20 and 10 respectively in this paper.

Suppose that a feature point set is composed of moving object, if a higher similarity between a block and the feature point set, a higher saliency the block will be, and therefore S_i can be seen as saliency of block i. After S_i regulated to $[0,255]$, Motion saliency map can be obtained.

Experimental Results

In order to verify the effectiveness and robustness of this algorithm, a variety of standard video were tested for experiment. The proposed motion segmentation algorithm is tested and compared with related work in the experiment.

Coastguard test videos are multiple moving objects scenes in camera motion conditions. The test sequence contains the global motion of camera and also contains the relative background of the vessel as well as local motion of undulation. In Figure 1, the first line is the coastguard video motion segmentation results. In the first 50 frames, there are two moving objects, through unsupervised clustering algorithm in this paper, the movement is divided into three categories. Where the first category of feature points is represented by a circle and the other two were treated with cross and square. As can be seen from Figure 1 (a) (b) the three different kinds of motion consist of moving objects and background are successful decomposition, leaving only one moving object after the 50 frames. By the pretreatment of mean shift and unsupervised clustering of this paper, feature points of moving objects and background are divided into two groups. As shown in Figure 1 (c) (d), moving objects are successful separated from background. As a conclusion from the experiment, the trajectory clustering algorithm proposed in this paper can adaptive determine the number of clusters under the scenes of motion camera.

Coastguard			
a	b	c	d

Walking man			
e	f	g	h

Figure 1　　Experimental　results of motion segmentation
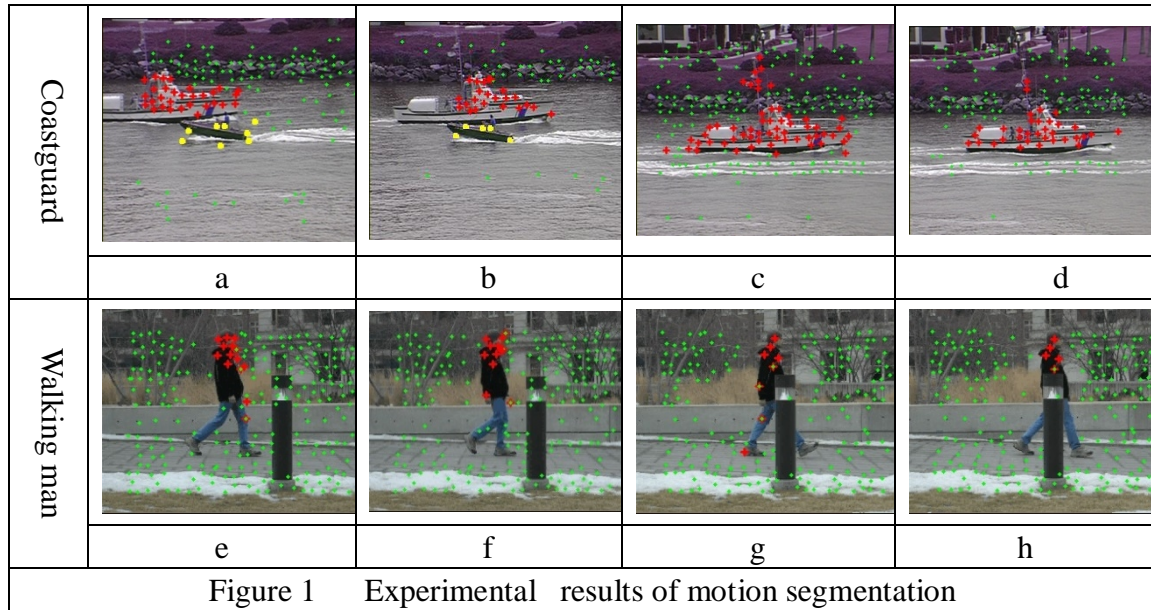
The second line (walking man) of figure 1 is the video clips of a pedestrian tracking by camera. As the character motion is non rigid, so the traditional motion parameter model couldn't work on Character motion modeling and the methods in [6,9] are invalid. The second line of figure 1 is the result of the test videos of this paper's method. When q is set to 5, the trajectory is divided into two categories. As can be seen from Figure 1 (e, f, g, h), the feature points belonging to the people were clustered into one group. This is because the method using trajectory clustering technology of this paper, by tracking feature points and generate the trajectory, the feature point segmentation is obtained. As the results shown in the experiment, this method has good adaptability to the rigid and non-rigid motion. However, while the following two frames have a covering will led to the failure of the feature point matching. But as a result of the pre classification according to the path length, this method still can be very good to different motion segmentation. From the test results of four consecutive frames of figure 1, the robustness of the method can be achieved.

Conclusions

In this paper, we propose an efficient visual saliency detection method based on trajectory clustering. Using the correspondence between the feature points of consecutive frames, feature points trajectories are obtained. Then, according to the trajectory information, feature points are classified through a special clustering algorithm. The new method obtain motion information according to the corresponding relationship of feature points, it contains less noise and does not depend on the specific motion model and use the motion information of consecutive frames . Also, it does not necessarily require any adjacent frames are continuous and stable enough. Experimental results on different videos demonstrate the utility and performance of the proposed approach.

Acknowledgments

References

[1] Itti, L, Automatic foveation for video compression using a neurobiological model of visual attention, IEEE Transactions on Image Processing,2004,13(10): 1304 – 1318

[2] Fred Stentiford,Attention-based similarity,Pattern Recognition,2007,40(3):771-783

[3] Harris C G,Stephens M J . A combined corrter and edge detector [A]In : Proceedings Fourth Alvey Vision Conference[C]. Manchester,UK . 1988:147-l51

[4] Bradski, G.R, Pisarevsky.V. Application in calibration, stereo, segmentation, tracking, gesture, face and object recognition.[C] IEEE Conference on computer vision and pattern recognition, SC, USA, 2000: 796-797

[5] Salari V, Sethi I K. Feather point correspondence in the presence of occlusion[J]IEEE Trans PAMI, 1990,12(1):87-91

[6]Shrinivas J. Pundlik，Real-Time Motion Segmentation of Sparse Feature Points at Any Speed, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,2008,38(3):731-742

[7]Xiaogang Wang, Kinh Tieu, "Learning Semantic Scene Models by Trajectory Analysis," in Proceedings of European Conference on Computer Vision (ECCV) 2006

[8]Gianluca Antonini ,Jean Philippe Thiran, Counting Pedestrians in Video Sequences Using Trajectory Clustering, IEEE Transactions on Circuits and Systems for Video Technology, 2006:16(8): 1008-1020

[9] O. Le Meur1, D. Thoreau1, A spatio-temporal model of the selective human visual attention[C] IEEE International Conference on Image Processing(ICIP 2005),Genova, Sept, 2005:1188-1191